An Overview of Particle Sampling Bias

Robert V. Edwards
Case Western Reserve University
Case Center for Complex Flow Measurements
Cleveland, Ohio 44106

and

James F. Meyers
NASA - Langley Research Center
Hampton, Virginia 23665

Second International Symposium on
Applications of Laser Anemometry to
Fluid Mechanics
July 2-4, 1984
Lisbon, Portugal

An Overview of Particle Sampling Bias

Robert V. Edwards
Case Western Reserve University
Chemical Engineering Department
Cleveland, Ohio 44106
Case Center for Complex Flow Measurements

and

James F. Meyers
NASA - Langley Research Center
Mail Stop 235A
Hampton, Virginia 23665

Abstract

The complex relation between particle arrival statistics and the interarrival statistics is explored. It is known that the mean interarrival time given an initial velocity is generally not the inverse of the mean rate corresponding to that velocity. Necessary conditions for the measurement of the conditional rate are given.

Introduction

The problem to be dealt with is that in a sparsely seeded flow, the probability of recording a velocity v depends on:

(1) the probability of the velocity v appearing in the measurement volume.

(2) the probability of a particle arriving at the measurement volume when the velocity is v.

(3) the probability of detecting a velocity from a particle of velocity v even if it passes through the volume.

(4) the probability of recording a measurement.

In general, all of the latter effects depend on the velocity itself. For example, for a flow with uniform spatial seeding, the higher speeds will carry more particles per unit time through the measurement region than will the slower speeds. Also, for various reasons, the electronics are less likely to record a higher speed particle than a lower speed one {Durao and Whitelaw (1979), Durao, Laker and Whitelaw (1980)}. If one records a histogram of measured velocities at a fixed point, $p_m(v)$, it is related to the Eulerian velocity distribution by an equation of the form

$$P_m(v) \quad N\, p(v)\, h(v)$$

where N is the total number of measurements, and $h(v)$ is the conditional probability of recording a measurement **if** the velocity is v. Another interpretation of the term $h(v)$ is the **relative measurement rate** for the velocity v. It has been written {Edwards and Jensen (1983)}

$$h(v) \quad \frac{r_m(v)}{r_m}$$

where $r_m(v)$ is the measurement rate for velocity v and < > denotes expected value.

Under the best of circumstances, $h(v)$ is difficult to compute {Buchhave (1975)}. Worse, the assumptions one has to make to compute $h(v)$ are often not valid. Many times the seeding particle density is not uniform. Under these circumstances, any correction applied to eliminate the effect of $h(v)$ may actually bias the data more. Problems of this type were noted by {Hoesel and Rodi (1977)}. They pointed out that residence time correction was inappropriate unless the particle density was uniform. They went on to suggest that if the particle density could not be assumed to be uniform, then residence time weighting is not appropriate. One should attempt to measure $h(v)$ by looking at the particle interarrival times.

Statistics of Particle Interarrival Time

If the flow is steady **and** the seeding is uniform the probability of the time t between particles is of the form

$$P_I(\ ) \quad r(v)\, e^{\ r(v)}$$

where $r(v)$ is the mean arrival rate corresponding to the velocity v. This follows readily from the fact that the volumetric distribution of the particles follows a Poisson distribution. In general $r(v)$ can be written

$$r(v) \quad |v|\, A(\vec{v})$$

where is the particle density and $A(\vec{v})$ is the effective measurement volume cross section. It has the dimensions of area and is positive definite. The cross section can be a function of flow angle.

For this steady flow, the average interarrival time τ_v is given by

$$_v \quad (r(v))^{\;1}$$

For a turbulent flow, the situation is more complex since the relevant probability is the conditional probability of an interarrival time τ_v **if** the initial velocity is v. It is not possible to give a global computable expression for the function, however the asymptotic behavior is known. At times small compared to the microscale τ_M, the expected rate remains *r(v)* and at times larger compared to the microscale, the flow is uncorrelated with v and thus the expected rate becomes $<r>$. An approximate form for $P_I(r,v)$ the probability of an interarrival time given the initial velocity v can be derived using the methods described in {Edwards and Jensen (1983)}. However, for illustrative purposes, a much simpler form can be used, i.e.,

$$P_I(\ ,v) \quad \begin{cases} \dfrac{\exp(\quad r(v)\ \ )}{_I} & (0 \qquad _M) \\[2ex] \dfrac{\exp(\quad (r(v) \qquad r\ \ )\ _M)}{_I} \exp(\qquad r \quad) & (\ _M \qquad) \end{cases}$$

$$_I \quad \frac{1}{r(v)} \qquad \frac{1}{r} \quad \frac{1}{r(v)} \exp(\quad r(v)\ \ _M)$$

This form ignores the expected variance in rate for long times, but that effect is demonstrably small. Figure 1 shows some forms of this function for various mean rates and initial velocities. Using this formula, the mean interarrival time, given the initial velocity v is given by

$$_v \quad r \quad \frac{1 \quad ((r(v) \quad _M \quad 1) \quad \left(\dfrac{r(v)}{r}\right)^2 \quad (1 \qquad r \qquad _M))\, e^{\ r(v)\quad _M}}{_I\, r^{\,2}(v)}$$

Clearly, the mean interarrival time is not simply the inverse of the rate. For instance, when the expected arrival rate *r(v)* approaches zero, the mean interarrival time does not approach infinity. This reflects the fact that the velocity only stays near any given value for a time on the order of τ_M. The asymptotes are:

$$(1) \quad r(v) \quad {}_M \quad 1, \quad {}_v \quad \frac{1}{r} \, \frac{1 \quad r \quad {}_M \quad \frac{(\; r \quad {}_M)^2}{2}}{(1 \quad r \quad {}_M)}$$

$$(2) \quad r(v) \quad {}_M \quad 1, \quad {}_v \quad \frac{1}{r(v)}$$

Only in the second, high particle density, case does τ_v appear to look like the inverse of $r(v)$. For the more typical, low particle density case, the mean interarrival time is a weak function of v. See Table 1 for a more detailed examination of the variation of τ_v with rate.

The weak dependence of $<\tau_v>$ on r_v is apparent in the measurements of {Durao, Laker and Whitelaw (1980)}. In their measurements, they included a non-negligible reset time, but this complication does not negate the above explanation of the measurement results.

Effect of Instrumentation

As has been noted in many articles {Durao, Laker and Whitelaw (1980), Edwards and Jensen (1983), and Meyers and Clemmons (1979)} the particle arrival rate is not the same as the measurement rate. The use of Bragg cells, the existence of counter reset times, etc., can all alter the recorded statistics. This result is a measurement rate for a velocity v, $r_m(v)$, that is different from $r(v)$. In the rest of this presentation, the rate $r_m(v)$ will be used. A general derivation for the statistics of any instrumentation set up cannot be given, however as before a few general statements can be made:

(1) The expected rate for times short compared to τ_M, will be $r_m(v)$.

(2) The expected rate for times large compared to τ_M will be $<r_m>$.

Recall that $r_m(v)$ is a conditional probability so that the times referred to above mean the times after the occurrence of v in the flow.

The difference between $r_m(v)$ and $r(v)$ can be illustrated by examining some data taken by Stevenson, Thompson, Bremmer and Roesler. In that study, they make velocity measurements in a turbulent flow while varying the effective particle density. The data collection system had a reset time (dead time) that was shorter than the flow correlation time. As the particle density increased past the point where the product of the arrival rate and the reset time exceeded one (saturated), the measured means changed. See figure 2. This clearly indicated a change of the measured statistics from those of the particle arrival statistics.

{Edwards and Jensen (1983)} derived an approximate form for $h_m(v)$ for a system with a reset time T. viz.

$$h_m(v) \quad \frac{r(v)}{r} \frac{(1 \quad r \quad T\ )}{(1 \quad r \quad (T \quad R_T) \quad r(v)\, R_T)}$$

where $R_T \quad \int_0^T R(\ )\, d$, R( ) is the velocity autocorrelation function, and is a normalization constant. Again one can gain some insight into the expected behavior by considering the asymptotic behavior of this expression.

$T << T_c = R_{T=\infty}$, The Integral Correlation Time

Under this condition, $RT \quad T$. Then

$$h_m(v) \quad \frac{1}{\ } \frac{r(v)}{r} \frac{(1 \quad r \quad T\ )}{(1 \quad r(v)\, T\ )}$$

(1) If $r(v)T >> 1$, many particles arrive during the reset time. Then $h_m(v) = 1$, there is no effect of the particle arrival statistics.

(2) If $r(v)T << 1$, few particles per reset time, then $h(v) \quad \frac{r(v)}{r}$

The particle arrival statistics are apparent in the measurement statistics. The solid line in Figure 2 reflects this change in the statistics as the arrival rate is varied.

$T >> T_c$

This corresponds to a reset time longer than the flow correlation time. Then

$$R_T \quad R \qquad T_c$$

At this asymptote,

$$h_m(v) \quad \frac{1}{}\,\frac{r(v)}{r}\,\frac{(1 \quad r \quad T\,)}{(1 \quad T_c\,(r \qquad r\,))}$$

For either $r(v)T >> 1$ or $r(v)T << 1$, the particle arrival statistics are reflected in the measurement rate.

For reset times longer than the flow integral time scale, the measured rates behavior is vastly different from the behavior of the mean interarrival time. With a large reset time, the interarrival probability function only contains information about the mean flow. Using the assumptions used earlier,

$$P_I(\ ) \quad \begin{matrix} r \quad e^{\ \ r} & \text{for} \quad T \\ 0 & \text{otherwise} \end{matrix}$$

With this probability function, $\ _v/<\ > = 1$. This is clearly shown in {Durao, Laker and Whitelaw (1980)} results for a reset time larger than the flow correlation time.

The meaning of the above is that although a system with a reset time larger than the flow correlation time will have a mean interarrival time that is independent of the initial velocity, the mean arrival rate is **not** independent of the original velocity. To understand how this can happen, one must realize that the interarrival time distribution is a reflection of the conditional probability of **another** measurement if a measurement of v is obtained. On the other hand, the measured rate, $r_m(v)$ is a reflection of the probability of getting the initial measurement of v.

Measurement of $r_m(v)$

As was shown above, one cannot measure $r_m(v)$ (and thus $h(v)$) by measuring the mean particle interarrival times. Even worse, if there is a dead time larger than the flow correlation, one cannot determine $r_m(v)$ from the interarrival time statistics. However it is conceivable that estimates for $r_m(v)$ can be obtained by examining the data rate for times that are small compared to the microscale.

For a steady flow, it is easy to measure $r_m(v)$. One simply picks a time interval Δt and for successive non-overlapping intervals, measure the number of measurements one gets. The mean number of measurements divided by the time Δt is an estimate of the rate $r_m(v)$. Roughly, if N_v is the total number of measurements, the relative error in the estimated rate is $(N_v)^{-1/2}$. This follows if each measurement is independent. For some circumstances such as a saturated detector, {Edwards and Jensen (1983)} or multiple measurements of the same particle, each measurement is not independent and thus the relative error will be larger.

For a turbulent flow, the velocity is not constant and the probability of getting two or more measurements of the same velocity with a laser anemometer is very small. However if one makes a histogram of the measured velocities using a finite number (say K) of non-overlapping ranges, one can arrange the divisions so that each range contains at least two measurements. The rate estimates can be performed for each interval of the histogram.

Selection of Histogram Intervals

In the worse case, an estimate of the worse fractional change of the rate over a velocity range Δv_H, is $\Delta v_H / v_M$, where v_M is the average velocity in the interval. This assumes a rate proportional to the velocity as exemplified by McLaughlin and Tiederman's one dimensional models. Most other models give a weaker dependence on velocity. Let Δv_R be the range of measured velocities, or 4 standard deviations, whichever is larger. The change in rate in each interval compared to the change in rate across the entire velocity range is roughly $\Delta v_H / \Delta v_R$.

For a real data set with a finite number of measurements, the above considerations place contradictory requirements on the selection of Δv_H, the width of the histogram intervals. Accurate estimation of rate in each range demands a large number of sample measurements and thus as large a Δv as possible. On the other hand, accurate resolution of the change of rate across the measured range, Δr_R, requires a small Δv_H. We do not know of a procedure for optimizing Δv, but experimentally have settled on $\Delta v_H / \Delta r_R = 1/9$.

A detailed exposition of one procedure for estimating $r_m(v)$ and thus of deriving the true Eulerian velocity probability distribution, $p(v)$, from the measured distribution is given in a companion paper by J. Meyers.

Sample and Hold

Processing the recorded data by a sample and hold scheme is exactly equivalent to estimating integrals of the velocity by a forward step integration algorithm. One holds the previous velocity value until a new one is obtained. {Dimotakis (6)} had proposed a backward step algorithm for use in the situation of many measurements per flow correlation time. There is no essential difference in the results obtained for a forward or backwards integration scheme. If the average of the measurement per flow correlation time is small, then the approximation to integration fails. However, when the measurements per flow correlation time is large, the approximate integration schemes are good approximations to the continuous integrals {Edwards and Jensen (1983)}. When these latter conditions are obtained, sample and hold processing or the method suggested by Dimotakis is clearly the best method to use. The particle statistics are avoided.

Conclusions

The measured arrival statistics for laser anemometers in sparsely seeded flow is indeed complex. No theory can adequately predict these statistics as many uncontrollable and unmeasurable variables in the system can influence the statistics to an important degree. Therefore unless the rather specialized and rare conditions of many measurements per flow correlation times are obtained, one should **not** use any of the previously proposed *corrections*.

The mean interarrival time between measurements given an initial velocity v is related to the mean measurement rate in a complex manner. In some cases where the measurement system cannot record measurements in a time shorter than the flow correlation time, the mean measured interarrival time is a constant. This effect can be used to get an order of magnitude estimate of the flow correlation time.

If the measurement system is capable of measuring particles separated in time by less than the flow microscale time, it is possible to measure the required correction function. This can be done by measuring the measurement **rate** for each velocity in time ranges that are small compared to the correlation time of the flow.

References

1. Durao, D. F. G.; and Whitelaw, J. H.: *Relationship Between Velocity and Signal Quality in Laser-Doppler Anemometry*, J. Phys. E: Sci. Instrum., 12, 47 (1979).

2. Durao, D. F. G.; Laker, J.; and Whitelaw, J. H.: *Bias Effects in Laser Doppler Anemometry*, J. Phys. E: Sci. Instrum., 13, 442 (1980).

3. Edwards, R. V.; and Jensen, A. S.: *Particle-Sampling Statistics in Laser Anemometers: Sample-and-Hold and Saturable Systems*, J. Fluid Mech., 133, 397 (1983).

4. Buchhave, P.: *Biasing Errors in Individual Particle Measurements with the LDA Counter Signal Processor*, Proc. LDA Symp. Copenhagen 258 (1975).

5. Hoesel, W.; and Rodi, W.: *New Biasing Elimination Method for Laser-Doppler Velocimeter Counter Processing*, Rev. Sci. Instrum., 48, no. 7, 910 (1977).

6. Dimotakis, P.: *Single-Scattering Particle Laser Doppler Measurements of Turbulence*, AGARD Conference No. 193, Applications of Non-intrusive Instrumentation in Fluid Flow Research paper 10.

7. Meyers, J. F.; and Clemmons, J. I., Jr: *Processing Laser Velocimeter High-Speed Burst Counter Data*, **Laser Velocimetry and Particle Sizing**, eds., H.D. Thompson and W.H. Stevenson, Hemisphere Publishing Co. p. 300 (1979).

8. Stevenson, W. H.; Thompson, H. D.; Bremmer, R.; and Roesler, T.: *Laser Velocimeter Measurements in Turbulent and Mixing Flows - Part II*, Air Force Tech. Rep. AFAPL-TR-79-2009, Part II.

$\frac{r_v}{r}$	r_M 0.1 r_v	r_M 1 r_v	r_M 10 r_v
0.0	1.005	1.25	5.55
0.1	1.004	1.23	4.56
0.5	1.002	1.13	1.96
1.0	1.000	1.00	1.00
1.5	0.998	0.87	0.67
2.0	0.995	0.74	0.50

Table 1.- Variation of $_v$/< > with rate for various mean rates.

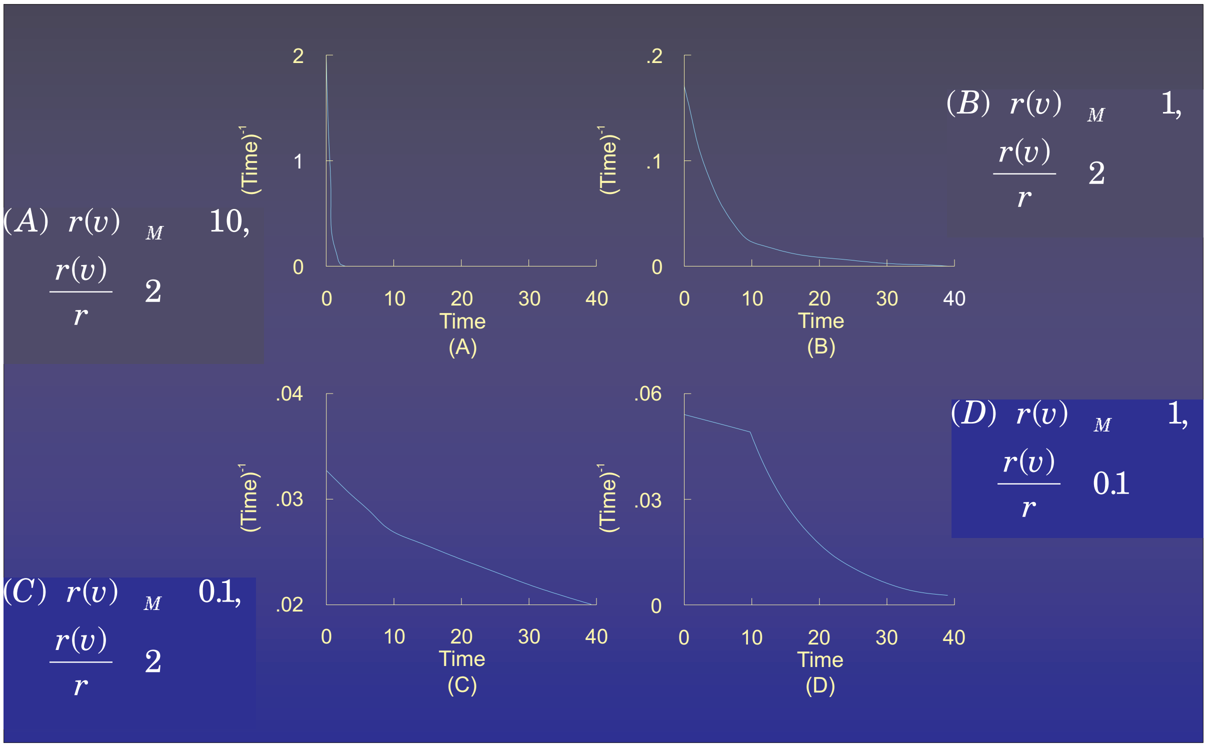

Figure 1.- Approximate conditional probabilities for the interarrival time given the initial velocity as a function of apparent particle concentration. The flow correlation time in each figure is 10.

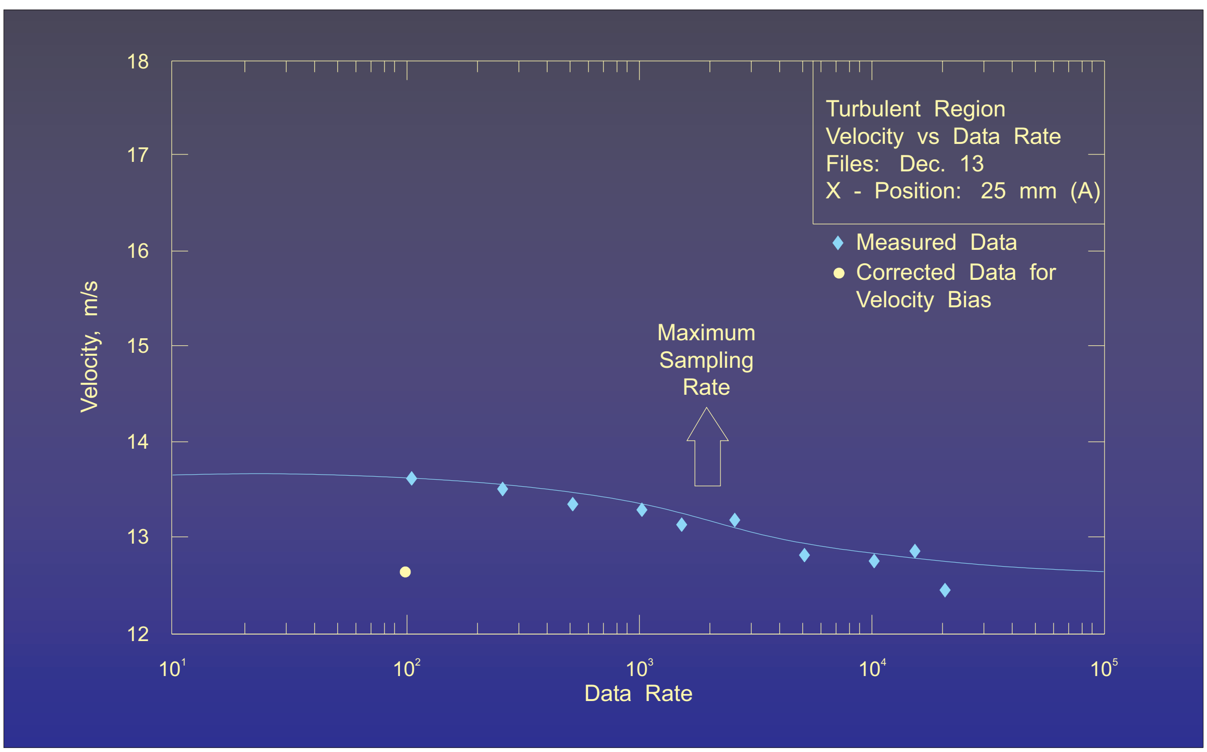

Figure 2.- Velocity vs Data Rate at Location A.